0. Preliminary remarks

Credit and other codes

This demo was developped by Laurent Pueyo (STScI) in *Mathematica* in order to take advantage of an easy on the eye user interface (and because the STScI ALICE pipeline is in *Mathematica*).

We are aware *Mathematica* is not commonly used by astronomers. Publicly available PCA based PSF subtraction routines can be found at the following adresses:

- GPI pipeline (IDL) : http://www.planetimager.org/index.php?option=com_content&view=ar-ticle&id=21&Itemid=36

- D. Mawet and ESO group pipeline (IDL): http://www.sc.eso.org/~dmawet/Dimitri-Mawet/IDL_PCA_pipeline.html

- Zurich Pynpoint pipeline (Python): https://pypi.python.org/pypi/PynPoint-exoplanet -AMNH/NYU S4 pipeline (Matlab): http://www.amnh.org/our-research/physical-sciences/astrophysics/research/project-1640/supporting-material

The data for this demo is public (HST) however condiderable work has been carried out by Remi Soummer, Elodie Choquet, Brendan Hagan, and Glenn Schneider to gerenarte these high quality cubes of coronagraph images. Should you play with them and find interretsing scientific objects, please refrain from serious archive stalking and agressive publishing and pleased contact Remi Soummer, (PI of the ALICE program) and Glenn Schneider (PI of the LAPLACE program) for civilized coordination.

• How to run this tutorial to high contrast PCA.

a) Read introductory text, which should give you some background on what the section of this session should be about.

b) When there is a mathematica input command, eg something that looks like this:

```
Brazil = 1;
Germany = 7;
Germany - Brazil
```

б

then click somewhere on that cell and press Shift + Enter. This will execute the command for you. After a command that will open a GUI you will see the comment ``this is what it should look like" in order to make sure that the command ran properly. Sometimes you will have to move the sliders on the GUI before it looks exactly like the screenshots in the notebook.

All reductions are done in real time. Calcualtions that are not intantaneous will be denoted beforehand by the :

Serious work

tag. All other executable cells are simply variable initialization. And should run instantly.

c) Look at the GUI, play with the buttons, try to make sense out of it, talk with your peers, ask questions.

d) When you are done with the GUI and want to move to the next step go in the bottom left corner of the GUI window and click on the button "Click Here to Close". Failure to do so will not clear the memory, varaibles will overwrite on top of each other and this demo might crash. In that even please go the the menu bar, click on ``Evaluation" then on ``Quit Kernel". This will clean everything up nicely and you will be able to start that part of the session from scratch. Since

e) Please ask questions is something is not clear.

1. PSF subtraction

This routine takes two contiguous NICMOS rolls that have been aligned and scaled and subtracts them one to another.

Serious work

Get["/ssw/Tue/Part3/SaganPSFSub.m"];

Play with the min and max display to make the positive / negagative images of the display appear. What can we say about the presence of a companion?

This is what it should look like



Classical roll to roll PSF Subtraction

2. PCA

First we need to set up the reduction parameters

The parameters to split the images in zones are

- Nphi number of azimuthal regions (or of pie slices)
- r0 part of the PSF core that we just simply mask out because the data is too crappy
- Δr width of the annulii (as a fraction of the total width of the image)

```
Nphi = 3;
r0 = 0.1;
\Delta r = 0.3;
Nr = Floor\left[\frac{1}{\Delta r}\right];
```

The following command is just an illustration that shows how the image is split in zones.

Serious work

Get["/ssw/Tue/Part3/SaganMakeZones.m"];

Running the PCA

We have 6 on the source we are looking at with NICMOS, three per roll.

```
NumberOfFiles = 6;
```

We apply a PCA on each zone and generate two reduced images, one for each roll in the NICMOS data. We have over 400 images in the PSF library but we choose here to only use the 200 that are the most correlated.

Setting the parameters:

```
Nphi = 3;

r0 = 0.1;

\Delta r = 0.3;

Nr = Floor\left[\frac{1}{\Delta r}\right];

Ncor = 200;
```

Now we run the PCA. What can we say about the reduced images?

Serious work

```
Get["/ssw/Tue/Part3/SaganRunKLIP.m"];
```

```
1
2
3
1
2
3
1
2
3
4
5
6
```





PCA Subtraction So what is the point ?... a little bit of speckle statistics

Everybody says that LOCI/PCA make planets easier to see in the data, but this time around I could sort of see the point source using PSF subtraction. What is all the fuss about? Think about that statement for a second and what might be not completely right.

If PCA indeed helps with the detection problem then one should be able to measure the imporvement. Here is an attempt to do so : by running the script below you will have a display of an histogram of the pixel counts in an annulus right inside of where the point source is (top) and in an annulus that includes the point source.

Discuss what we can conlcude about the statitical sinificance of the detection.

Serious work

Get["/ssw/Tue/Part3/SaganRunKLIPWithStats.m"];

```
1
2
3
1
2
3
1
2
3
4
5
6
{17.876488, Null}
```

However at this point we should have a discussion regarding what these histograms means and how one could use them to derive SNRs and confidence intervals. I expect everybody to argue vehemently about SNRs.

This is what it should look like



Printed by Mathematica for Students

3. Photometry

Assessing the problem

We can see in the tail of the histograms that the flux is getting eaten up by the PCA algorithm. Really how bad is it?

Using the same parameters as before :

```
Ncor = 200;
Nphi = 3;
r0 = 0.1;
\Delta r = 0.3;
Nr = Floor \left[\frac{1}{\Delta r}\right];
NumberOfFiles = 6;
Serious work
Get["/ssw/Tue/Part3//SaganAperturePhotometryNoCalibration.m"];
1
2
3
1
2
3
1
2
3
4
5
```

6



Solution 1 a) : creating a dataset with only the fake candidate

In the case of PCA we are simply projecting the image on the dominant modes of the PSF library. Since the science images is composed of speckles + astrophysical signal, and since projecting onto principal components is a linear operator, then one can in principle calibrate the algorithm throughput based on the set of fake images.

Using the same parameters as before :

```
Ncor = 200;
Nphi = 3;
r0 = 0.1;
\Delta r = 0.3;
Nr = Floor \left[\frac{1}{\Delta r}\right];
NumberOfFiles = 6;
```

Telling the algorithm roughly where the companion is so the fake ones can be injected ~ at the same place and sees the same PCA modes.

```
CandiatePosX1 = 66;
CandiatePosY1 = 39;
CandiatePosX2 = 62;
CandiatePosY2 = 27;
```

Running the reduction

Serious work

Get["/ssw/Tue/Part3/SaganAperturePhotometryCalibration.m"];

- _____

- б

- -

- б



The algorithm throughput measured with the fake candiates looks consistent with the throughput of the actual candidate. However when normalizing the latter by the former we

find photometric estimates that still widely varies with the number of PCA modes. This is due to two phenomena :

-a) the location of the fake candidate does not exactly correspond to the location of the actual candidate: they do not exactly see the same eigen-images.

-b) right now we are using a zone of~500 pixels over which the noise is highly correlated and we are calculating the PCA basis using the 200 most correlated images. This is very degenerate and we are very sensitive to the numerical noise in the higher order modes.

My claim is that the dominant effect for this dataset is b). Take some time to think of reasons why.

Solution 1 b) : creating a dataset with only the fake candidate and be smart about the ratio

of number of PSFs vs size of the zone

Reducing the number of modes

```
Ncor = 50;
Nphi = 3;
r0 = 0.1;
\Delta r = 0.3;
Nr = Floor \left[\frac{1}{\Delta r}\right];
```

NumberOfFiles = 6;

Telling the algorithm roughly where the companion is so the fake ones can be injected~ at the same place and sees the same PCA modes.

```
CandiatePosX1 = 66;
CandiatePosY1 = 39;
CandiatePosX2 = 62;
CandiatePosY2 = 27;
```

Serious work

Get["/ssw/Tue/Part3/SaganAperturePhotometryCalibration.m"];

```
{23.234904, Null}
```

This does not work very well, why?

Making the zone larger

```
Ncor = 200;
Nphi = 1;
r0 = 0.3;
\Delta r = 0.9;
Nr = Floor \left[\frac{1}{\Delta r}\right];
NumberOfFiles = 6;
```

Telling the algorithm roughly where the companion is so the fake ones can be injected ~ at the same place and sees the same PCA modes.

```
CandiatePosX1 = 66;
CandiatePosY1 = 39;
CandiatePosX2 = 62;
CandiatePosY2 = 27;
```

Serious work

Get["/ssw/Tue/Part3/SaganAperturePhotometryCalibration.m"];

This works well for the photometry, why?

However this does not tell us where the candiate is in the focal plane.

Solution 2 : Use forward modelling

Discussion (this will be covered in my talk tomorrow)

At this point we DO NOT inject fake candiate in the data. Instead we use a template PSF (usually normalized to the stellar flux, but this is not done here) is order to solve the following problem :

$$(\tilde{x}_{s}, \tilde{y}_{s}, \tilde{f}_{s}) = \underset{(x_{s}, y_{s}, f_{s})}{\operatorname{arg\,min}}$$

$$\left\| F - f_{S} \left(S - \sum_{q=1}^{K_{klip}} \langle S(\cdot - x_{S}, \cdot - y_{S}), Z_{q}^{KL} \rangle_{\mathcal{C}} Z_{q}^{KL} \right) \right\|_{\mathcal{F}}^{2},$$

$$(6)$$

Where:

- (\tilde {x_S}, \tilde {y_S}) correspond to the coordinates of the candidate in the focal plane (astrometry)

 $- tilde {f_S}$ to the relative flux of the the

candiate with respect to the flux of the template PSF.

This template is usually normalized to the stellar flux. However we did not thoroughly carry out this calibration step here. As a conquence the numbers that you will see only correspond roughly to the contrast of this source.

After some, not so tedious, algebra one can find that the astrometry is given by

$$(\tilde{x}_{s}, \tilde{y}_{s}) = \arg\max_{(x_{s}, y_{s})} \{ C_{(x_{s}, y_{s})}(F, S, K_{klip}) \}.$$
(7)

with

$$C_{(x_s, y_s)}(F, S, K_{klip}) = \langle F, S(x - x_S, y - y_S) \rangle_{\mathcal{F}} - \sum_{q=1}^{K_{klip}} \langle S(x - x_S, y - y_S), Z_q^{KL} \rangle_{\mathcal{C}} \langle F, Z_q^{KL} \rangle_{\mathcal{F}}, \quad (8)$$

and the photometry is given by:

$$f_{s} = C_{(\tilde{x}_{s}, \tilde{y}_{s})}(F, S, K_{klip}) \frac{C_{(\tilde{x}_{s}, \tilde{y}_{s})}(F, S, K_{klip})}{||S - \sum_{q=1}^{K_{klip}} \langle S(\tilde{x}_{s} - x_{S}, \tilde{y}_{s} - y_{S}), Z_{q}^{KL} \rangle_{\mathcal{C}} Z_{q}^{KL} ||_{\mathcal{F}}^{2}}$$
(9)

After some, seriously tedious this time, algebra these equations can betranslated into a piece of code heavily relying on Fourier Tansforms to calculate the various inner products and correlations. You can try this code below. The source code for this process is quite barbaric (e.g not well commented). Should you want to use this for your reseach in the future and something does not make sense feel free to contact pueyo@stsci.edu or any of the helpers who should be familiar with this method.

With small zones and plenty of reference PSFs

We still need to assume that we roughly know where the candidate is :

```
CandiatePosX1 = 66;
CandiatePosY1 = 39;
CandiatePosX2 = 62;
CandiatePosY2 = 27;
```

Using the same parameters as before

```
Ncor = 200;
Nphi = 3;
r0 = 0.1;
\Delta r = 0.3;
Nr = Floor \left[\frac{1}{\Delta r}\right];
NumberOfFiles = 6;
zonetest = 8;
(*Please ignore this variable but do not change anything here otherwise
 everything will crash and burn in a glorious fire of scientifc desperation*)
Serious work
Get["/ssw/Tue/Part3/SaganForward.m"];
1
2
3
1
2
3
4
```

- 5
- 6



Note that while the simple fake sources method diverged right away, the photometry is now stable over the first 100 modes. Why does it diverge? (attendees with whom I have already talked about this are not allowed to show off).

With small zones and less reference PSFs

We still need to assume that we roughly know where the candiate is :

```
CandiatePosX1 = 66;
CandiatePosY1 = 39;
CandiatePosX2 = 62;
CandiatePosY2 = 27;
```

Using the same parameters as before with less PSFs.

```
Ncor = 50;

Nphi = 3;

r0 = 0.1;

\Delta r = 0.3;

Nr = Floor \left[\frac{1}{\Delta r}\right];

NumberOfFiles = 6;

zonetest = 8;

(*Please ignore this variable but do not change anything here otherwise

everything will crash and burn in a glorious fire of scientifc desperation*)
```

```
Serious work
```

```
Get["/ssw/Tue/Part3/SaganForward.m"];
```

{3.724725, Null}

With Large zones

We still need to assume that we roughly know where the candiate is :

CandiatePosX1 = 66; CandiatePosY1 = 39; CandiatePosX2 = 62; CandiatePosY2 = 27;

Now the size of the zone is the full image:

```
Ncor = 200;

Nphi = 1;

r0 = 0.3;

\Delta r = 0.9;

Nr = Floor \left[\frac{1}{\Delta r}\right];

NumberOfFiles = 6;

zonetest = 1;

(*Please ignore this variable but do not change anything here otherwise

everything will crash and burn in a glorious fire of scientifc desperation*)

Serious work
```

Get["~/Desktop/SaganPCAPackage/SaganForward.m"];

{13.614170, Null}

Final remarks

Looking at these examples provides a good picture of where the trade off are.We should discuss them. Q & A :

Why the little wiggles in the astrometry and photometry? Think before reading what is below.

- equations 6 to 9 assume that the residual noise is gaussian, if for some reason it is not, then they do not work ``perfectly". The little wiggles that we see in astrometry and photometry are due to the fact that the Gaussianity condition is not rigorously true.

What would break this method? Think before reading what is below.

- the refrence cube that we are using DOES NOT HAVE ANY ASTROPHYSICAL SIGNAL. This is because we are using other stars observed by Hubble. This is NOT true for ADI/IFU data and some more work is needed in order to use this estimator in these cases (see tomorrow' s lecture).

Is it all we need to worry about for astrometry and photometry? Think before reading what is below

- in the present example the pixel coodinates are known within~1/20 th of a pixel (~3.5 mas in the case of NICMOS). This IS NOT THE ASTROMETRIC uncertainty as there is a slew of calibrations to carry out based on these numbers in order to get to the relative position of the companion with respect to the star (see tomorrow' s lecture).

4. Disk detection

Now that we have seen that PCA work pretty well on an entiere NICMOS image we can look for extended souces. We do so using the following geomety, where again the entiere zone is the image.

```
Nphi = 1;
r0 = 0.1;
\Delta r = 0.9;
Nr = Floor \left[\frac{1}{\Delta r}\right];
NumberOfFiles = 6;
And we start by being agressive and use plenty of PSFs
```

Ncor = 300;

```
Serious work
```

```
Get["/ssw/Tue/Part3/SaganRunKLIPDisk.m"];
```

{51.485072, Null}

This is what it should look like



Something funky is happenning close in. We should try to be less agressive and use less PSF in the reference library.

Ncor = 100;Nphi = 1; r0 = 0.1; $\Delta r = 0.9;$ Nr = Floor $\left[\frac{1}{\Delta r}\right]$; NumberOfFiles = 6; Serious work Get["/ssw/Tue/Part3/SaganRunKLIPDisk.m"]; {17.638344, Null}

Indeed there is a disk!

Try to play with the parameters to make the disk apperas as nicely as with ACS.

To make things a little harder I have flipped the image (and yes than candidate we obsessed about for the past hour is a background source).



Note that our FOV is~6", so we are exploring the inner edge of the disk.

The outer halo has been supressed by the PCA but can be recovered using classical PSF subtraction : both methods are highly complementary in this case.